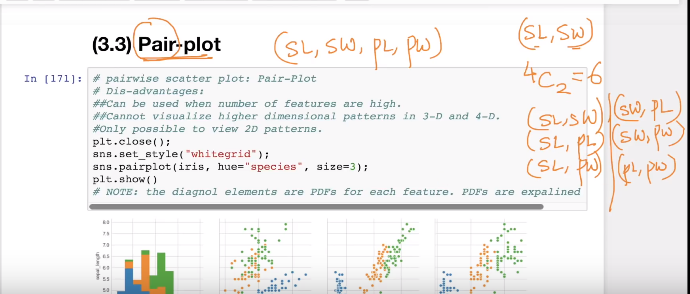
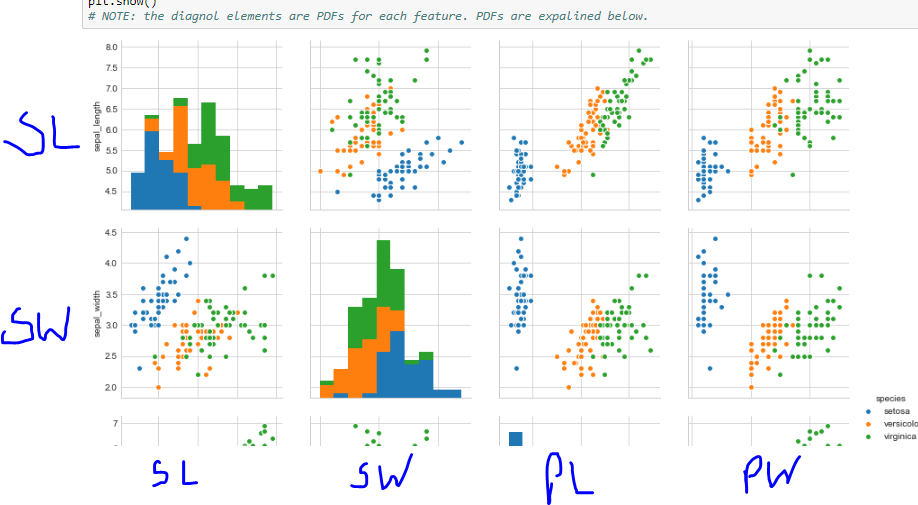
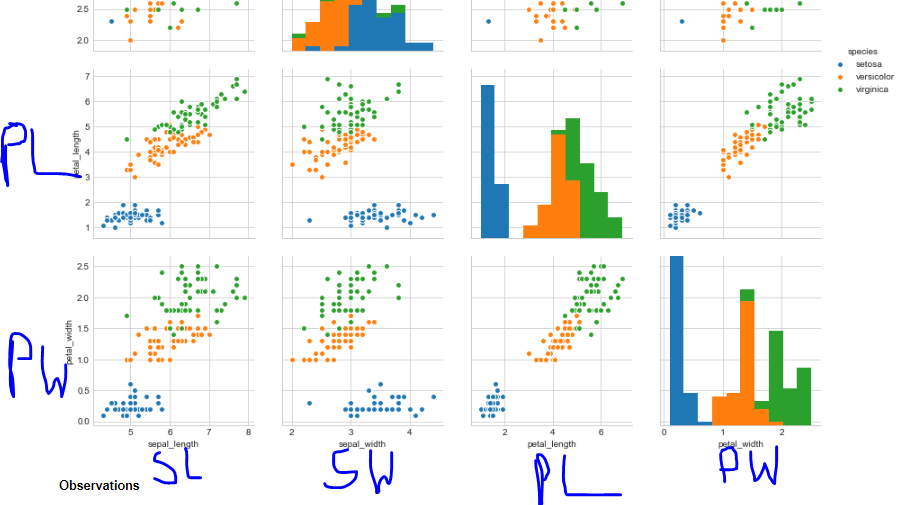
It becomes very difficult to visualize 4-d or more dimensions data, so in this case we use pair plots, in which we plot each feature with every other feature to find relation between them. For ex: for Iris data set there are 4 features and it’s become very hard to visualize them together, so we make pair plot as (SL, SW), (SL, PL), (SL, PW), (SW, PL), (SW, PW), (PL, PW), that means for 4 features there will be 6 plots ie 4C2.



Below figures show pair plot of iris data sets, as you can see in below figures the graph above diagonal are simply replication of graph below diagonal, just replacing X-axis with Y-axis and vice versa, so there are total 4 figures.





As you can see in above figures the plot of (SL & PL) and (SL & PW) are separating setosa iris with other two iris much better than (SL & SW) .

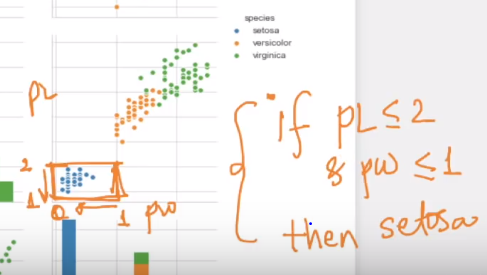


Since PL and PW combining with SL generating good separation so we can look for (PL & PW) plot

So looking at (PL & PW) plot mentioned below we can create basic model to classify setosa with other two as:

If PL <= 2 and PW <= 1

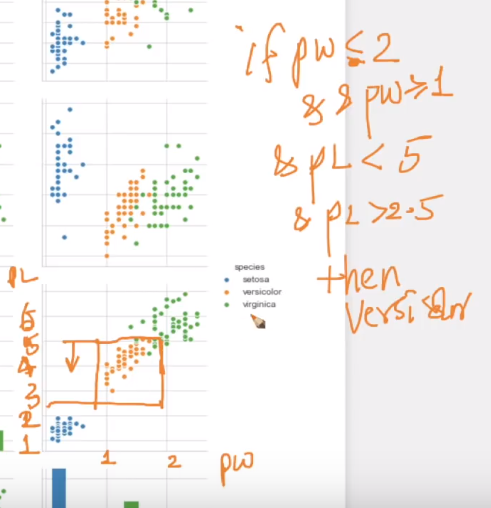
Then it’s setosa;



And we can also create model to classify versicolor as:

If PW <=2 and PW >=1 and PL < 5 and PL > 2.5

Then it’s versicolor



**Observations**

1. petal\_length and petal\_width are the most useful features to identify various flower types.
2. While Setosa can be easily identified (linearly seperable), Virnica and Versicolor have some overlap (almost linearly seperable).
3. We can find "lines" and "if-else" conditions to build a simple model to classify the flower types.

**Limitation of Pair plots**

Suppose if there are large number of features like 100, 200 then it will be very difficult to bring out observation from them, so in such datasets with large features we use **PCA, t-SNE.**